Introduction

There has been a growing trend in democratic societies to *engage* the public in activities and decisions that have in the past been the preserve of experts and/or designated authorities. This is particularly the case with regards making *policy* decisions, ranging from setting agendas or priorities in services to informing novel legislation (e.g. Bovaird, 2007; O'Faircheallaigh, 2010; Owens, 2000; Putnam, 2000). The scientific domain appears to be a particularly fertile one for engagement (e.g. Pieczka & Escobar, 2013; Pytlik-Zillig & Tomkins, 2013), with relatively recent roots (notwithstanding other longer-term and subtle trends) in the controversy over the genetic modification of food and crops, where the attempted introduction of this technology and its products has been essentially seen as a failure, due in large part to technology and policy advancing without adequate consideration of public values and desires (e.g. Horlick-Jones, Walls, Rowe, Pidgeon, Poortinga, Murdock & O'Riordan, 2007). Indeed, *public engagement* is such a compelling zeitgeist that the main issue in research and practice seems to be how or when it should be done, rather than *if* it should be done. Here we discuss a number of dilemmas with involving the public in policy decisions specifically, although most (if not all) of our following analysis is liable to be applicable to public engagement in any domain whether this leads to the formal development of policy or not.

In this paper, we take as our starting point the definition of Rowe and Frewer (2005), who use the term *public engagement* to encompass all activities in which authorities interact with the public (and its representatives) on some issue, and retain the term *public participation* for activities in which information flows to-and-fro *between* these parties. The interactive nature of participation allows it to be differentiated from *public communication* and *public consultation*, in which there is a one-way flow of information only (i.e. from public to authorities or vice versa). Our interest here is in *participation* in particular, which is essentially seen as a superior alternative to simple *public communication*, for reasons discussed next.

The justification for *participation*, in brief, is that traditional processes of democratic governance and knowledge regulation are inadequate - exacerbated by a decline in (or diaspora of) trust in policy makers, scientists and 'the establishment' more generally (Jones, 2015). In correcting a slide towards 'bowling alone' (Putnam, 2000), the public – or its representatives – is invoked as a legitimized and licensed contributor to the policy making process and mobilized as a conversationalist and commentator on all sorts of issues, from the permissibility of novel technologies to the appropriateness of solutions to environmental problems (Hansen, 2010; O'Doherty & Einsiedel, 2013; Chilvers & Kearnes, 2015). The public(s) – it is conjectured – are purveyors of lay wisdom not possessed by policy makers, which includes local and specific knowledge, but also insight into the popular mind and consciousness – which is extremely important, given that many of us live in

relatively liberal democratic societies, in which some degree of public support is often necessary to enable policies to be enacted. Aside from this, democratic theorists would argue that it is simply *right* to increase public involvement in societal affairs in order to increase learning, political efficacy, societal cohesion, and so on (e.g. Dewey, 1927; Dryzek, 2000; Fishkin, 2011; Gutmann & Thompson, 2009).

Accepting public participation to be a desirable activity (as many, though not all, do), the issue is then how to enact the process. Numerous variations on a limited number of themes exist. However, most approaches are group-based and one-off (although enduring action can be achieved through committee membership), manipulating relatively small numbers of lay people to confront a problem/ agenda/ policy and to think about and deliberate solutions (Rowe & Frewer, 2005). The big question of what participation actually has achieved remains largely unanswered, mainly because evaluation is rare – being a difficult, expensive, controversial, and theoretically uncertain process (e.g. Rowe & Frewer, 2000; 2004), that is susceptible to eliciting awkward truths that might threaten the industry that now supplies it (Lee, 2015). In the absence of much in the way of evidence, it might be worth stepping back to appraise the likelihood – or indeed, possibility – of public participation effectiveness.

This paper will suggest that there are a number of dilemmas which, together, may work *against* the achievement of effective participation in policy-making processes. These dilemmas have been largely informed by our experience as both academics in the area as well as practitioners of evaluation over many projects (in the UK and Europe) over a significant number of years. Although we support our assertions with appropriate references where we can, it is important to note that formal supporting evidence is often hard to come by, in large part because of a lack of relevant case studies or other empirical evidence in the peer-reviewed literature (and even in the grey literature). There are liable to be several reasons for this, from the fact that much of the work in conducting and evaluating participation is done by practitioners and consultants who lack the motivation (and perhaps capacity) to publish their results, to the problem of what constitutes acceptable 'evidence' (in the eyes of editors and reviewers) in such a difficult-to-control area, to potential biases against publishing pieces that contest the worth of such a popular concept (an issue that at least one of our 'dilemmas' considers). As such, we invite the reader to accept our analysis on a level at which they are comfortable: we have stated six 'dilemmas' (we toyed with the term 'paradoxes'), although we would be equally happy for these objects-of-analysis to be considered six 'propositions' or 'hypotheses' instead. Our hope is that this analysis stimulates debate and perhaps may subsequently be used as a formal justification by others for publishing work or case studies that more directly address the dilemmas we have stated.

In the rest of the paper, these dilemmas will be stated, explained, and justified in turn. Following a discussion of the dilemmas, their implications as a whole will be considered and potential solutions to likely problems are suggested.

The Timing Dilemma

The more appropriate the timing of public participation, the less likely it will actually take place.

One issue discussed in the literature is *when* participation should take place. It has been argued that much participation has traditionally taken place essentially as an afterthought, or worse, as something of a 'public relations' exercise *after* a decision has actually been made on the topic of concern, with little scope for the public to do more than *affirm* that decision, and no real hope of providing actual influence. Consequently, there has been discussion of the need for 'upstream engagement' (Wilsdon & Willis, 2004), that is, for engagement (participation) to take place much earlier in a process when real influence is still a possibility. While we have sympathy with this position, we also think that there are a couple of significant reasons why such timing for participation is liable to be difficult.

First there arises the ontological matter of what, exactly, is the nature of the newly emergent issue. Given that such an issue – like an area of science – is novel, then it is likely to be fluid and ill-defined. What, then, are the boundaries of the issue? What does the issue involve, and what does it not involve? How does it differ from related issues? These are difficult questions which – at such an early stage – will be contested to a greater or lesser extent by the specialists/ experts/ scientists in that domain. And yet, in order to conduct participation, a coherent characterisation is required in order to *inform* public participants of the key elements/ aspects so that they are *then* able to indulge in relevant debate. The importance of such information is particularly great here as the public will have commensurately less information about the issue (e.g. nanotechnology) on which to base *any* sort of judgment or reasoning than they would have on a more familiar issue (e.g. nuclear power); indeed, many may not have even *heard* of the issue and will have no prior information at all on which to draw.

Second, and also related to the area's novelty, there arises the issue of agency: the new area is liable to be immediately populated by domain experts (e.g. scientists) of a particular type. Given the focus of these agents on developing and understanding the new issue, it is unlikely that their premier concern will be in discussing matters with the public (they are at the stage of needing to discuss matters with other high-level specialists). It is likely that specialists or practitioners of participation will thus arrive later on the scene, and that these will need significant support from the domain

experts before they are able to appropriately understand the nature of the new issue and the kind of questions that may be apt for public participation. Of course, difficult does not mean impossible, and we are aware of concerted attempts by a number of projects to engage the public early on topics such as nanotechnology and synthetic biology. One of us is, for example, currently involved in a project on synthetic biology, which, perhaps unsurprisingly, has seen rather more stakeholder participation (experts attempting to come to consensus on the nature of the new area of science) than public participation, with much of the latter best characterised as initial attempts to *communicate* key concepts than have *meaningful* two-way dialogues.

In short, the more advanced is an issue, the greater the amount of coherent information that will have accumulated, and thus the easier it will be to *informationally resource* public participation and provide a clear set of questions to be debated. This is particularly important when real *influence* is intended to be offered to the public: experts and policy makers – we suggest – are less likely to accept public recommendations when the very issue they are engaging with is fluid and open to disagreement (being novel, with contested definitions), than when it is rather closed and mature, with stakeholders of different persuasions at least able to agree on what are the critical facts that are needed by the public participants in order to begin their dialogue. Hence, although we might agree that participation is best suited early on in a process, ontological and agency matters mean that participation is more likely to occur the later an issue is along its trajectory of development when many of the key decisions will have already been made. Put another way, if participation is best when conducted early in a process (one evaluation criterion for effectiveness e.g. Rowe & Frewer, 2000), then most cases will prove suboptimal. Furthermore, it is unclear whether the results from any early stage participations that are conducted will have much potency or credibility in driving their still-nebulous problem area (and we await any evidence at all to this effect).

The Relevance Dilemma

The greater the need for public participation (specifically, deliberation to resolve issues between competing parties), the less likely it will take place.

Public participation invariably involves *deliberation*. In simple terms, deliberation involves having an interactive conversation in which information (facts, opinions, values) is exchanged between two or more parties. In public participation, the information exchange occurs between members of the public, as well as between the public and other more knowledgeable stakeholders (experts, scientists, policy makers, etc.). The purpose of deliberation is to enhance the understanding of *all* of those involved and can – when the object of a piece of participation is to develop some sort of

advice, agenda, recommendation, or other output – involve *negotiation* and attempts to reach a greater degree of consensus (we say 'greater degree' as full consensus is unlikely). Importantly, for this to be possible, *all* parties must be open to enlightenment, including the 'more knowledgeable' parties: should one 'side' be closed, then the exercise essentially diminishes to a piece of communication (a one-way process in which information flows from a rigid source to a malleable receiver e.g. Rowe & Frewer, 2005), in which the public members would be expected to learn or change their minds but not the stakeholders (etc.). Indeed, this latter case would typify the sort of tokenistic public participation that is frequently condemned.

It can be argued that the need for/ value of deliberation increases in line with the extent of 'controversy' related to an issue. That is, when an issue is uncontentious, deliberation appears to offer little beyond confirming the acceptability of an option. Thus, when the 'facts' about a particular issue are well known and understood (knowledge is high and there is little uncertainty surrounding the 'opinions' of experts/ stakeholders), and the *values* of stakeholders and the public are essentially aligned ('essentially', as there is no topic on which all will agree) – or perhaps the values themselves are more static than dynamic – then engagement would not seem a particularly valuable approach. That is, communication of unknown facts may be all that is necessary, with at most a tentative check that the public have no great concerns about the issue at hand (i.e. public communication and public consultation may suffice instead of public participation - Rowe & Frewer, 2005).

The greater the controversy, the more there is to discuss, and indeed, the greater is the need to negotiate and reach a tolerable degree of consensus so that some form of decision (if required) can be made, or at least, areas and reasons for lack of consensus be identified. Greater controversy may emerge when facts are contested (opinions on what are the facts, and which facts are valid and which are not), and when values differ (an issue is seen as variably positive or negative by different people, in terms of physical or emotional costs or benefits). Differences in *values*, as opposed to *certainty of knowledge*, would seem to be the key factor here. Consider where facts are well known, such as in engaging the public on a transportation plan: all may accept the knowledge claims of the experts (who may also be in agreement amongst themselves), yet controversy may arise because different options cost or benefit different people differently.

However, the state of knowledge is not an insignificant factor in determining controversy potential, as low knowledge/ high uncertainty environments are open for people to *differentially interpret and use* facts to support their values/ positions. As an aside, consider for example, Fischer, Van Dijk, de Jonge, Rowe and Frewer (2013) who provided neutral information on nanotechnology to subjects in an experiment, and found that some became more positive about the technology while others became more negative. That is, the participants appeared to be interpreting the information in line

with their existing values. Indeed, this is a widely recognised feature of human psychology: people tend to search for information to support their existing position and attempt to interpret information in a supportive way, while dismissing contrary evidence as invalid and its advocates as unreliable and biased (the general phenomenon is often termed the *Confirmation Bias* – see, e.g. Plous, 1993). As such, people are not good natural scientists, in the logical positivist sense, as they ever seek to confirm their hypotheses about the world, rather than disconfirm them (as nicely demonstrated in the Wason Selection Task, e.g. Wason, 1966). For a number of fascinating examples in how people with firm beliefs can make the most incredible mental contortions to get emerging evidence to fit in with their existing beliefs, see Christine Garwood's (2007) book on the Flat Earth idea and its various protagonists.

Ideally, then, participation might theoretically be most helpful when it is suspected that the public may have differing opinions to those of the authorities responsible for a problem (if not, why ask them in the first place?). We posit, however, that public participation in science policy is more likely to be seen as an option by responsible authorities when controversy is low than when expected to be high. In latter cases, various vested interests – particularly fear of ceding power to a party that might come to different conclusions to the sponsor – are liable to pre-empt engagement in the first place. Indeed, the motives of those sponsoring public participation cannot be assumed lily-white. For example, the politicization of knowledge more generally (or in other words its manipulation and habitual reduction to what Henderson (2012) calls 'policy-informed evidence') is an explicit example of the vulnerability of science to co-option as an instrument of political brokerage and gamesmanship (hence why not public participation too, in which the evidence comes from a less reliable source?).

Additionally, even if a participation project *is* initiated in a controversial area, it is likely such a project will fail – either to reach a conclusion (because of difficulties resolving disputes, thanks to the psychology behind the confirmation bias), or prove influential. We have personal experience of such an instance. Several years ago the UK government set about holding a major public participation on the controversial issue of genetically modified (GM) foods and crops (we were part of a team that won the contract to evaluate the event). However, this event was soon abandoned when a representative of an anti-GM movement, involved in the steering committee, very publicly withdrew from the project, claiming that the framing of the project was biased to ensure a pro-GM public answer would be attained (a position that it suggested was desired by the government sponsors). Now, whether true or not, this exemplifies how difficult it can be – when at least one party has strong and immovable views – to allow a process to proceed that might undermine one of the main protagonist's positions.

In short, this dilemma suggests that public participation in science policy will generally not be allowed for scenarios where it might be of greatest relevance, as it needs to be commissioned by an authority with power related to the problem, which is unlikely to wish to do so unless it is relatively certain of receiving a positive outcome. In many cases, then, the public participation that is conducted will largely be an (expensive) irrelevance.

The Representation Dilemma

As the intensity (and arguably quality) of a participation process increases, the representativeness of its output is likely to decrease, undermining its credibility.

Public participation, as noted earlier, tends to involve relatively small numbers of the public (Rowe & Frewer, 2005) involved in intensive activities. Small numbers are necessary because participation processes invariably involve dialogue, and this is simply not possible (nor affordable) with large numbers. However, the lack of numbers calls into question the 'representativeness' of the sample of the wider population, and yet such representativeness is critical - and this may be exacerbated if continued engagement is required over a number of events or period of time (e.g. see Mihaly, 2010). Indeed, 'representativeness' is an important common criterion in evaluating participation quality (e.g. Rowe & Frewer, 2004), while the concept is particularly significant for sponsoring authorities, as it enables them to counter criticisms of biased process membership and justify more widely the validity of the process output and its credibility. Indeed, our personal experience in discussing such matters with senior national and European decision makers is that 'representativeness' is often their main concern, and that this often lies at the heart of scepticisms about the ultimate utility of participation activities.

Advocates of participation *do* tend to take the concept of representativeness seriously, in the sense that they attempt to recruit a sample that broadly reflects the typical characteristics of the sampled population (e.g. having equal gendered participation). They recognise the numbers problem, and counter that participation aims are largely qualitative – intended to find out about the nature and nuance of public opinions – than quantitative (differentiating it from the intentions of surveys), and that their samples provide opinions that are 'indicative' as opposed to formally 'representative'.

Put another way, there are questions of diagnostic reliability where what is claimed as either success or failure in an engagement is an assertion predicated on limited numerical evidence. For example, given the significant cost implications of most engagement activity and what some allege as the associated difficulty in easily recruiting 'sufficient numbers' of public participants on a voluntary (or non-incentivized basis), numbers of recruited public participants are habitually modest (no more

than approximately twenty to thirty). Where exit-poll data then reflects the attitudes of this number as either positive or negative, for instance in terms of how they think their input will be used in the policy oeuvre, the impact of this claim is necessarily diminutive. Credibility regarding the representativeness of the public voice is furthermore compromised by sampling strategies that *infer social inclusivity through topographical breadth*. For example, where engagement is operationalized on a UK level, commonly the case across a range of urban conurbations (e.g. London, Birmingham, Edinburgh, Cardiff), geographical range is erroneously conflated with public representativeness. And while participants may be profiled on the basis of social determinants like gender, ethnicity and age, no account is really made of the spectrum and inter-fusing of different kinds of 'local' lay knowledge that will inform what they have to say. Unsurprisingly therefore, and in tune with what has been our own experience, relatively homogenous groups are assembled both within and between different engagement locales. Of course, organising these kinds of events is a far easier undertaking in the city context than in peri-urban and rural settings. Fundamentally, however, the validity attributed to public engagement outputs is most stretched and at risk of repudiation where devoid of any kind of statistical significance.

Regardless of initial selection, the process of participation may also introduce and exacerbate biases. Thus, where public engagement through deliberative dialogue processes is intended to increase the representativeness and therefore efficacy of debate by recruiting a more plural and diverse body of perspectives, it is concurrently vulnerable to the exchange of ideas becoming blocked. This may be due to an inability to reach consensus where the attitudes and outlooks of multiple discussants are polarised and influenced by social backgrounds and life experiences that are unlike and perhaps even antagonistic (again, see discussion above on confirmation bias). Discussants may also be distinguished by variance in their 'dialogical capital' – manifesting unequal levels of confidence and/or skill in argumentation – and may be inhibited or otherwise likely to dominate discussion. The kinds of power relations at play within multi-actor discussions may therefore mean that engagement may be only superficially 'representative' where discussants may be included yet not fully integrated or 'licensed to participate'. Furthermore, where the moderation of dialogue, by for instance a workshop convenor or facilitator, is weak (and/or overly interventionist), the potential for equal and representative dialogue further recedes while the threat of bias augments. Accordingly, while heterogeneity in the membership of engagement activities is seen to be desirable it is also highly problematic, where the reality of a differentiated public is a less than harmonized continuum of public opinion and ostensibly, therefore, less usable conclusions.

This then is where engagement advocates and organisers promote online and digital spaces of engagement as a panacea to the problem of numbers, where deliberative dialogue is potentially

enacted on an escalated even universal scale. The promise of a global conversation however is yet to fully materialise, and despite huge advances in the context of digitally mediated social connectivity, the public/policy online interface remains at best clunky, often characterised by stilted, inefficient, asynchronous, difficult-to-understand communications (e.g. Rowe & Gammack, 2004). This may be partially to do with an issue of propriety but also ownership, control and regulation by governing authorities and cognate risks of non-containment and corruptibility, certainly where involving sensitive or ethically contentious topics, where online communication platforms facilitate the diffusive and unbounded spread of knowledge (Watermeyer, 2010; 2013). So then, the dilemma of representativeness for participation will likely persist. Geographical breadth may be achieved and so too depth in the level of well regulated (and choreographed) face-to-face dialogue, but the proportion of participants will remain, certainly at a time of extended austerity in the distribution of public funds and the kind of which supports such activity, low and probably self-selected or made homogenous by their motivations to be involved (something they regularly do; on the basis of a financial incentive; or having vested interest in the topic at hand). For all these reasons, credibility of outputs will suffer, particularly as stakeholders who are antithetical to any options or conclusions expressed by the public participants will ever be able to challenge the legitimacy of their representation.

The Evaluation Dilemma

The more a participation exercise needs to be evaluated (because of potential difficulties), the less likely it is that it will be adequately evaluated.

It is clear that there are many participation events ongoing throughout the world at any time, although relatively few of these will have been evaluated in any formal sense (although Rowe & Frewer, 2004, published an early review of participation exercises, this considered ones reported in the academic literature only; we are unaware of any more recent review). We posit from experience that most of the relatively small, inexpensive events will lack any form of evaluation *at all*. However, there does appear to be a trend for larger, better-funded events to have at least *some* form of evaluation (that is, *some* larger events - but certainly not *all* or even *most*). In the UK, for example, *Sciencewise*, which describes itself as 'the UK's national centre for public dialogue in policy making involving science and technology issues' (www.sciencewise-erc.org.uk) – and which until 2016 was funded by the UK Government's Department for Business Innovation and Skills - *has* co-funded many large engagement exercises over the last few years, and in doing so it has required these to *also* commission independent evaluations. We are not aware of any other major organisation that

oversees national or international participation events that has a similar requirement, although it may be the case that certain funders (e.g. governmental agencies in some countries) do indeed have such a policy. What is clear is that evaluation is not ubiquitous, and even when this is done it is not likely done with any great degree of rigour. There are many possible reasons for this, including expense, knowledge of how to conduct evaluations, and even *motivation* (e.g. Rowe, Horlick-Jones, Walls & Pidgeon, 2005).

Put another way, the participation events that *are* evaluated are likely to be characterised by sponsors or advisors that have the extra funding to afford evaluation; have a greater knowledge about the concept of participation and of evaluation; and have a higher motivation to evaluate – perhaps at least partly consequent on that greater knowledge of processes (encouraging a conviction that any event they oversee *will* prove of acceptable quality). Hence, the *very presence* of evaluation is likely to be a signal of a well-resourced and considered event. That is, the very presence of evaluation should – as with the looming necessity of an exam – work to enhance the quality and relevance of what is done. To some extent, this is good: the mere existence of evaluation should improve the enactment of participation. The problem is, in the absence of evaluation, anything goes - and indeed, anything can be claimed.

There is a further confounding factor here, in that many of those evaluations that are conducted tend to focus on *formative* evaluation as opposed to *summative* evaluation. That is, an evaluation rationale may be predicated upon the demonstration of 'due process' and 'good practice' and for the purpose of exercising transparency concerning what was done and its effects. Assessing the qualities of such practice and its outcomes at the end of the event would count as summative evaluation. Evaluation of participation may also, however, be rationalised in less instrumental terms and instead on the basis of a pedagogical contribution to formative learning and as an iteration of critical reflection that ameliorates participation practice. It is in this context, especially, that evaluation occurs, not as is so often the case in an *ex-post* format, but as a cradle-to-grave dimension that feeds into, enriches and potentially modifies its operation. This, in our specific experience of *Sciencewise* evaluations (as an example), is the dominant *modus operandi* - yet it is one that also complicates the role and function of the evaluator and his/her relationship with those undertaking and sponsoring the participation process. For one, the critical distance of the evaluator may unwittingly erode where s/he becomes integrated into the running of an exercise. This places a different set of expectations on the evaluator who thus becomes more accomplice than commentator and whose brief is the facilitation of an event's success and less the determination of such. This kind of reassignment or reimagining of the evaluator role is at once well intentioned - where continuous evaluation provides for live and responsive learning – while at the same time it

opens up the prospect for bias and evaluative blindness where the evaluator is too close, immersed and complicit with the evaluated process. This may also introduce an unrealistic and unreasonable expectation of the evaluator, where the cost association of evaluation tends to be calculated at only ten per cent of an engagement. This is but one pressure on evaluators – others will be discussed shortly.

Considering all participation events as a whole, then, this leads us to expect that the evaluated subset of events will be atypical of the universal set and more likely to include 'good' events than 'bad' ones (including events that may have started as 'bad' but then were enhanced by formative evaluation). From this, we expect positive evaluations to be the norm (to the extent that the evaluations themselves are valid – an issue we turn to next), which may give an unrealistic impression that participation works in cases where evaluated, leading to the faulty extension that non-evaluated events are also likely to be good (and the conclusion: participation is indeed, by default, intrinsically, good).

The Criticism Dilemma

The more critical are evaluators and evaluations – and therefore, the greater the necessity they imply for restitutional action – the less impact they will have.

We previously posited that evaluated events are more likely to be 'good' than un-evaluated events. This is liable to be the case both in fact and as confirmed by the evaluations themselves. However, we think that evaluation reports are likely to be biased on the upside (i.e. not completely valid), as evaluators find themselves under various pressures to be positive (as discussed below), further clouding how successful public participation as a generic activity truly is.

Perhaps the greatest threat to credible and/or 'honest' evaluation of participation – where what is taken to be 'honest' is confirmed by the evaluator's adherence to agreed criteria that embody what works well and what works less well – would be the intercession of those with vested interest in and aversion to anything less than the broadcasting of positive outcomes and impacts. In this scenario, where what is reported within evaluative findings is perceived to be 'negative' or with negative connotations or implications, stakeholders, specifically those most proximal to and 'involved' within the evaluation process (e.g. a commissioning agent or body) may be inclined to intervene in the selection of what is reported. Stakeholders of this kind may encourage or tacitly enforce changes in the presentation of evaluative findings so as to reflect an interpretation of events more akin to their own, typically benign and partisan, persuasions. Where the stakeholder is a paying client and where the evaluator's remuneration for services undertaken and potential for future contracting is

predicated on client satisfaction, an opportunity to contest or even reject such 'overtures' is limited. Where contestation does materialize between evaluator and client, this may be cause for significant additional labour; prolonged 'negotiation'; and even perhaps the total redaction of a report from public view. At root is an issue of who owns the evaluation and a struggle of compromise and concession, which in a worst case scenario may be the evaluator's lost integrity.

Where the client claims ownership and authority over what an evaluation report says, the (critical) agency of the evaluator not only enervates but the evaluation process itself becomes redundant or worse, fraudulent. Certainly, the confidence within which evaluative findings are interpreted as impartial, objective and underpinned by ethical and methodological rigour may wane. Furthermore, the authority of the evaluation report, frequently held as the *de facto* report of the participation exercise, dissipates where the client assumes editorial control; where analysis is corrupted by intervention by persons outwith the evaluation process; and where a client's expectation of evaluation is not as a process yielding critical insight and formative learning but as one delivering pre-ordained results.

Unfortunately, our wide practical experience of conducting evaluations suggests that the scenario just discussed is more frequent than not – and indeed, is the norm. Simply put, sponsors of public engagement are human, often working within a political environment. Contractors who deliver public engagement are also human. Unfortunately few humans take criticism well, particularly when such criticism is written in black and white and is available for superiors, peers, and competitors to inspect. This sensitivity seems to be especially the case for those we call *True Believers* (people almost religiously convinced of the rightness of participation) who have conducted a piece of participation, and who have *invested significant emotional and material resources* into running an event. Their response to even the gentlest critique can often be a denial of evaluation claims, and a questioning of the quality of the evidence supporting such claims. Equally unfortunately, because evaluation is difficult, and there is a lack of validated instruments and processes (see Rowe, Horlick-Jones & Walls, 2005), it can be difficult for evaluators to *prove* the correctness of their claims. The irony is that those opposing critical claims on the basis of lack of evidence will invariably cite even *less* worthy/ rigorous evidence to support their own (inevitable) claims of event success ('but everyone was smiling, and two people came up to me after to say how much they enjoyed it!'). Even worse, refutations can become very personal – attacking the evaluator as well as the evaluation. We have experienced several such attacks. In one (for the record, conducted in 2014), one of us provided an extremely carefully worded critique of the conduct of an event (in which no individual was identified as bearing any personal responsibility for anything) that – in reality – was a disaster. Indeed, this was widely recognised by a large number of supporting members of the overall project

in which this engagement took place, who were involved in the process in a number of ways (including as facilitators of break-out groups). In spite of this, the head of the (unduly) *mildly* criticised event reacted with fury, accused the evaluator of deliberate bias (although he was unable to indicate *a single place* in the evaluation report where such bias was evident, in spite of a number of requests for him to do so), and undoubtedly pressurised the project coordinator to refuse the report (unsuccessfully). In another case, noted in Rowe *et al* (2005), the evaluators were actually accused of industrial espionage in response to (again generally mild) criticism – clearly an *ad hominem* attack aimed to distract from the message of the evaluation. Thus: evaluators who are critical are invariably seen as biased people who have completely misunderstood a process, and hence critical evaluations can safely be dismissed as poor pieces of work that should be ignored (*viz*, buried in archives where they cannot be found). Of course, some humans are more enlightened and willing (often through gritted teeth) to contemplate critiques as valid and respond in a constructive way – though in our experience the proportion of such beings is *extremely* small.

The consequence of this is that mild assessments of generally well-conducted events may prove more visible and influential than more serious critiques of less well-done events, which may be amended, downgraded, and hidden. An extension to this problem is that critical evaluators not only lose influence on the evaluated process but – through reputational issues – face reduced influence in the wider community, as more malleable evaluators gain a positive reputation for doing inoffensive 'light touch' evaluations and gain commissions instead. The problem is exacerbated by the tendency of funders of participation events to also be the funders of their evaluation, and hence able to select amongst proposals from several evaluators that which appears most likely to give them the positive evaluations they desire. Indeed, we have personally been told on a number of occasions that we missed out on gaining commissions because our rigorous, concept-based evaluation approach 'scared' funders.

In short, official, published evaluations are likely to be biased to the upside, once again giving an unduly positive picture of the quality of public participation events, individually and as a whole.

The Impact Dilemma

The greater the focus of evaluations on the impact of participation events, the less informative (and more misleading) they will be.

Evaluations of participation in science policy debates increasingly demand a focus not only on *process* but *impact*. Indeed, because evaluations cost money, a focus on impacts may come at the cost of reducing time and scope to evaluate processes. It has been argued elsewhere that evaluation

of processes is essential and should be predominant because high impact from a badly conducted event (e.g. one with biased information and an unrepresentative public sample addressing ill-thought out questions) would ultimately be a failure, though a misdirected evaluation might record it as a success (e.g. Rowe & Frewer, 2000). A good example of this is the 'GM Nation?' event conducted in the UK in 2003-2004. This event suffered serious problems process-wise, not least in gaining participation from a highly partisan sample (whereas its explicit aim was to sample the 'grass roots' general public). It then concluded with an incredibly negative message on the perceived merits/risks of genetically modified foods that was inaccurately reported in the media as being the view of the general public, and which arguably made the subject a no-go area for the UK government for many years (e.g. Horlick-Jones, Walls, Rowe, Pidgeon, Poortinga & O'Riordan, 2006; Horlick-Jones, Walls, Rowe, Pidgeon, Poortinga, Murdock & O'Riordan, 2007; Pidgeon, Poortinga, Rowe, Horlick-Jones, Walls & O'Riordan, 2005). A focus on impact alone would lead to the conclusion that 'GM Nation?' was a huge success – but in reality, it was a terrible failure.

Process and impact should, we suggest, be seen much in the same way as the concepts of *reliability* and *validity* are seen in ascertaining the value of a measurement tool, that is, good process should be seen as a *necessary though not necessarily sufficient condition* for event success (in the same way that reliability is a necessary but not necessarily sufficient feature of a valid instrument). Thus, demonstration of good, acceptable process *must* come first so that the implications of any subsequent impact can then be judged as success or failure (and one would hope that a poorly conducted event would have *no* impact!). Unfortunately, it seems clear to us that in the search for evidence of impact, participation event sponsors are underplaying the issue of process quality, probably because they do not understand the importance of process issues but do understand the need to demonstrate impact.

So, what is the source of this new emphasis? Much of a move towards measuring the impact of participation (and engagement more generally) relates to the formalisation of impact – or scientists articulating the economic and societal impact of their research - as a mandatory feature of both competitive research funding and national research evaluation processes (Watermeyer, 2014). The ascent of auditing cultures in professional scientific and knowledge-generating communities demands that the cost-benefit of public participation/ engagement in science requires thorough justification and thus forms a core aspect of the evaluation process. However, evidencing impact causality in the context of participation is habitually difficult if not implausible. The impact(s) of participation may have a longer than shorter gestation period and certainly cultivate more slowly than the demands of rapid evaluative reporting will allow. Consequently, evaluators may be forced to only speculate upon future impacts or elucidate juvenile, highly localised or superficial impacts

such as those 'imagined' by stakeholders. Similar to the difficulties faced by researchers in forecasting (Chubb & Watermeyer, 2016) or historicising (Watermeyer, 2015a,b) the impacts of their research, the impact-visualization of engagement is impaired by claims that participation is not an iteration of - but conduit to – impact, and that even where attributed as impact, this is of a 'soft' or more nebulous variety (Watermeyer, 2014).

Overall, where evaluators are asked to report on the impacts of a given participation process a timetabled hiatus or suspension in the evaluation process would allow impact to develop and mature and therefore be more readily considered upon re-visitation of the activity. This would surely enable more credible and confident assessments of participation and release evaluators from having to second-guess impact's crystal ball. However, projects are rarely, if ever, designed to allow later assessments (most probably because of funding terminations).

In summary, an increased focus on impact evaluation at the expense of process evaluation is, we believe, misplaced, and will ultimately lead to evaluations that are less informative and more misleading.

Implications

So, what do these dilemmas actually imply? As a set, they make uncomfortable reading with significant negative connotations. In short, they suggest that public participation in science policy will not take place at the most appropriate times (early in a process) or for the most appropriate issues (those with greatest controversy), and that the participants recruited will often not be (or as importantly, *be seen to be*) suitably representative of the intended population. They suggest that, when participation does take place, the sort of participation that really needs to be evaluated will not be evaluated, while events that are most likely to be successful will be. However, even in the latter cases, there will be biases in the evaluations so that events generally appear more successful than they really are. This will result in a sampling bias, giving an unduly positive view of the merits of participation, which may be exacerbated by a focus on evaluating impacts at the expense of processes. These dilemmas suggest that the most rigorous evaluators will be punished and lose out to more pliant evaluators; that 'light touch' evaluations will predominate; that critical evaluations will be contested and dismissed (and indeed, prevented from being published or publicised), and that positive evaluations will proliferate. Consequently, evaluators will necessarily adapt to this reality (as we have to a significant degree) or be wiped from the evaluator gene pool. All these outcomes will also lead to bias in terms of the community's view of participation as a whole, reinforcing the extant views of the 'True Believers' and undermining the capacity for the field as a

whole to *learn*. This will ultimately lead to the detriment of the field, as requests by funders and sponsors of proof to justify expenditures (financial and otherwise) will continue to remain elusive and/or cause the triumph of artifice over honesty (Chubb and Watermeyer 2016), undermining the general case for *more* participation.

Indeed, a deeper understanding of what makes participation 'good' or not should not only lead to better *participation*, but should also bring into focus the question of whether *more* participation is actually desirable. There are a number of reasons why participation might fail, not all of which are down to the inept application of processes (though getting this across to sensitive sponsors and practitioners can be difficult!). One important possibility is that failure may be due to a mismatch between the process and the context (e.g. Rowe and Frewer, 2004, 2005) – that is, in some situations participation will almost always fail simply because it can never lead to any of the consequences typically hoped for, such as a meaningful opportunity for the public to truly influence a particular policy decision. Hence, we need to identify where and when participation is appropriate. Insufficient research has been conducted on when and what type of engagement is appropriate in relation to different types of decision (selective intervention). When is engagement not necessary? When is consultation or communication appropriate? When is co-production crucial and in what form? Without a proper understanding of participation quality, we lack the ability to resolve such issues.

Conclusion and Solutions

In this paper, we have touched on a number of dilemmas related to public participation in the area of policy making, particularly in the scientific context (although our analysis is liable to be transferable to participation in other domains and for other purposes). Our intent has not been to be comprehensive (indeed, there are liable to be many more dilemmas in the area), but to identify certain paradoxical situations that might have negative consequences. Addressing these issues will not guarantee the development of 'good' participation (but may be a start); good and appropriate participation will also require consideration of other important aspects, such as matching process to *purpose* (e.g. decision-making or feedback on services); taking account of the nature of the *place* in which participation occurs so that it is comfortable for participants; considering public views of the sponsors and their level of trust and using 'enlightened intermediaries' when necessary (e.g. Involve 2005) – and so on. But establishing whether to use participation, and if so, what particular format, implemented in what particular way, in what particular space, with what actors and intermediaries, should be based on *valid* and *trusted* learnings from *rigorous* and *fair* evaluations – which brings us

to our central conclusion: for public participation in science policy to achieve credibility – in the eyes of academics, policy makers, and stakeholders – practitioners need to take evaluation more seriously, and they have to become more honest about the positives, and particularly the negatives, of participation. They must accept that this is an emerging field and that there is still much to learn; they should look to social science for clues as to best practice; and they should embrace criticism and learn from negatives, rather than shutting down debate on evaluation (somewhat ironic, given that good dialogue is an aim of much participation). We suggest there are several steps that need to be taken to achieve this, as follows:

1. All participation events – no matter how small - should have *some* form of evaluation. Funders of events should insist on this, requiring practitioners to clearly state what their event aims to achieve and how they intend to find out whether those aims are achieved. Process evaluation is necessary and must not be excluded for impact evaluation alone.

2. Evaluators should, wherever possible, be independent from those sponsoring or running a participation event. In our experience, very few – if any – of those running an event are able to take a balanced view of their event: practitioners have too much invested (in terms of time, finances, reputation, and core beliefs) to see the full extent of negatives. Where it is not possible to have external evaluators, it may still be possible to seek external advice, such as in the development of a small user questionnaire (developed to ask about negatives as well as positives, and carefully designed to avoid leading questions and imbalanced scales) from methodological experts.

3. While stated sponsor aims are *necessary* criteria for evaluation and establishing event success, they are not *sufficient*. This is another reason why independent evaluation is needed. That is, sponsors often have only a nebulous understanding of what 'participation' is or should be, and little understanding at all about the issue of evaluation. In our wide experience, most sponsor aims tend to be so vague that they essentially cannot *fail* to be achieved, and most are what we term 'aims of intent', that is, 'we aim to engage with X members of the public, on issue Y, using method Z'. Within such aims there is no appreciation that process biases - in recruitment, in information comprehensiveness, comprehensibility and content (including different perspectives), and so on – need to be considered before any event can be declared a 'success'. External evaluators should be competent to bring such normative perspectives to the evaluation *without embargo* from the event sponsors/ actuators.

4. Evaluators should be appointed by an external, independent entity. It would be useful if such a national or international body could be established, although there are in many countries already public or non-profit making bodies that might serve such a role (e.g. in the UK, a body such as the National Coordinating Centre for Public Engagement, which supports a proactive culture of public engagement across UK universities). This is crucial. In our experience, bias begins at the selection of the evaluators, where those promising a 'light touch' are preferred to those promising rigour.

5. An evaluation report should stand, as produced. While we are not against sponsors/actuators seeing and providing some comments on a report – particularly to correct any clear errors of fact – our experience is that allowing sponsors/actuators any *power* to compel amendments is inimical. For example, we recently completed an evaluation (in 2016) in which the sponsor essentially told us to delete every single critical comment against the event being evaluated (indeed, this has happened to us on a number of occasions). What is an evaluator to do, when payment for services requires their report to be 'signed off' by those with vested interests? The result is that – as noted above - we have severe doubts as to the validity and balance of most evaluation reports we have seen (including some of our own). The model we prefer is for the evaluation document to be published as written (following opportunity for comment, entirely at the discretion of the evaluator to accept or otherwise, with payment for services guaranteed), and then for the sponsors/ actuators to have a first 'right of reply', to be published with the evaluation. In cases where this 'right' is taken up, the evaluators should be allowed final word, with their own right to respond to any issues raised. This model essentially follows that of critical pieces in journals that are published with commentary; it also follows an examination style process in which appeals are allowed and then accepted/dismissed. Indeed, the final word (or adjudication) could be published by an independent third body (as discussed in point 4), though that would perhaps only be necessary in a few occasions for major pieces of participation. Regardless, it must be made clear where evaluations truly are independent, and where evaluators labour under potential sanction (e.g. withholding of fees).

The above sequence is somewhat idealistic under current circumstances. An international body of public participation, perhaps populated by well-regarded experts and practitioners, perhaps with local 'chapters' to deal with regional proposals, might be of some value. Awarding events a blue riband for following all the five steps for rigorous evaluation noted above might temper an increase

in (probably relatively minor) negative evaluations, and provide the wider community of participation sponsors and users a hallmark of quality. Such an entity could also set up a database of accredited evaluators and reviewers, and develop acceptable rules or principles for evaluations to be signed-up to by all involved (funders, actuators, evaluators). For example, evaluations should be impersonal and balanced; they should be clear about the criteria used for evaluation and the methods/ instruments used for this purpose; they should commit to retaining data for future reference, and so on (in other words, they should accede to good social science practice). For sponsors, there should be an up-front commitment to publish the evaluation along with the participation report and to pay on time and without sanction for unpalatable reports. And so on.

Perhaps this ideal is some way off; regardless, it seems clear to us that, at present, there is a deep lack of honesty in the participation community. If True Believers really wish to promote participation, and see that public views have greater impact in areas in which they might not otherwise be found, then they need to accept rigorous evaluation; they need to accept the bad with the good. Only then will trust follow, and the true potential of public participation in policy become realisable.

References

Bovaird, T. 2007. Beyond Engagement and Participation: User and Community Coproduction of Public Services. *Public Administration Review* 67 (5): 846-860.

Chilvers, J. and Kearnes, M., eds. 2015. *Remaking participation: Science, environment and emerging publics*. London and New York: Routledge.

Chubb, J. and Watermeyer, R. 2016. Artifice or integrity in the marketization of research? Investigating the moral economy of impact statements within research funding proposals in the UK and Australia. *Studies in Higher Education*. DOI: 10.1080/03075079.2016.1144182.

Dewey, J. 1927. *The public and its problems*. New York: Holt.

Dryzek, JS. 2000. *Deliberative democracy and beyond: Liberals, critics, contestations*. New York: Oxford University Press.

Fischer, ARH., van Dijk, H., de Jonge, J., Rowe, G. and Frewer, LJ. 2013. The impact of information on attitudinal ambivalence: The case of nanotechnology in food production. *Public Understanding of Science* 22 (7): 817-831.

Fishkin, J.S. 2011. *When the people speak: Deliberative democracy and public consultation*. New York: Oxford University Press.

Garwood, C. 2007. Flat Earth: The history of an infamous idea. London: Macmillan.

Gutmann, A. and Thompson, D. 2009. *Why deliberative democracy?* New Jersey: Princeton University Press.

Hansen, J. 2010. *Biotechnology and public engagement in Europe*. London: Palgrave Macmillan.

Henderson, M. 2012. *The geek manifesto: Why science matters*. London: Corgi.

Horlick-Jones, T., Walls, J., Rowe, G., Pidgeon, N., Poortinga, W. and O'Riordan, T. 2006. On evaluating the GM Nation? public debate about the commercialisation of transgenic crops in Britain. *New Genetics and Society* 25 (3): 265-288.

Horlick-Jones, T., Walls, J., Rowe, G., Pidgeon, N., Poortinga, W., Murdock, G. and O'Riordan, T. 2007. *The GM Debate: Risk, Politics and Public Engagement*. London: Routledge.

INVOLVE. 2005. People and participation: How to put citizens at the heart of decision making. Eastleigh: INVOLVE.

Jones, O. 2015. *The Establishment: And how they get away with it*. London: Penguin.

Lee, C. 2015. *Do-it-yourself democracy: The rise of the public engagement industry*. New York: Oxford University Press.

Mihaly, M. 2010. Citizen participation in the making of environmental decisions – evolving obstacles and potential solutions through partnerships with experts and agents. *Pace Environmental Law Review* 27: 2009–2010.

O'Doherty, K. and Einsiedel, E., eds. 2013. *Public engagement and emerging technologies*. Vancouver: UBC Press.

O'Faircheallaigh, C. 2010. Public participation and environmental impact assessment: Purposes, implications, and lessons for public policy making. *Environmental Impact Assessment Review* 30 (10): 19-27.

Owens, S. 2000. 'Engaging the public': information and deliberation in environmental policy. *Environment and Planning A* 32: 1141-1148.

Pidgeon, NF., Poortinga, W., Rowe, G., Horlick-Jones, T., Walls, J. and O'Riordan, T. 2005. Using surveys in public participation processes for risk decision-making: The case of the 2003 British GM Nation? public debate, *Risk Analysis* 25 (2): 467-479.

Pieczka, M. and Escobar, O. 2013. Dialogue and science: Innovation in policy-making and the discourse of public engagement in the UK. *Science and Public Policy* 40 (1): 113–126.

Plous, S. 1993. The Psychology of Judgment and Decision Making. New York: McGraw-Hill.

Putnam, R. 2001. Bowling alone: *The collapse and revival of American community*. New York and London: Simon and Schuster.

Pytlik-Zillig, LM. and Tomkins, AJ. 2013. Public Engagement for Informing Science and Technology Policy: What Do We Know, What Do We Need to Know, and How Will We Get There? *Review of Policy Research* 28 (2): 197-217.

Rowe, G. and Frewer, LJ. 2005. A typology of public engagement mechanisms. *Science, Technology, & Human Values* 30 (2): 251-290.

Rowe, G. and Frewer, LJ. 2004. Evaluating public participation exercises: A research agenda, *Science, Technology, & Human Values* 29 (4): 512-556.

Rowe, G. and Frewer, LJ. 2000. Public participation methods: A framework for evaluation. *Science, Technology, & Human Values* 25 (1): 3-29.

Rowe, G. and Gammack, JG. 2004. Promise and perils of electronic public engagement. *Science and Public Policy* 31 (1): 39-54.

Rowe, G., Horlick-Jones, T., Walls, J., and Pidgeon, NF. 2005. Difficulties in evaluating public engagement initiatives: Reflections on an evaluation of the UK “GM Nation?” public debate. *Public Understanding of Science* 14 (4): 331-352.

Wason, PC. 1968. Reasoning about a rule. *Quarterly Journal of Experimental Psychology* 20 (3): 273-281.

Watermeyer, R. 2010. Social network science: Pedagogy, dialogue and deliberation. *Jcom*, 09(01) A04.

Watermeyer, R. 2013. Communication and the new media in Chadwick R, Levitt, M. and Shinckle, D. eds. *The right to know and the right not to know: 15 years on*. Cambridge: Cambridge Univeristy Press, 180-195.

Watermeyer, R. 2014. Impact in the REF: Issues and obstacles. *Studies in Higher Education*. DOI: 10.1080/03075079.2014.915303

Watermeyer, R. 2015a. Public intellectuals vs. new public management. The defeat of public engagement in Higher Education. *Studies in Higher Education*. DOI: 10.1080/03075079.2015.1034261

Watermeyer, R. 2015b. Lost in the third space: The impact of public engagement in higher education on academic identity, research practice and career progression. *European Journal of Higher Education* 5 (3): 331-347.

Wilsdon, J. and Willis, R. 2004. See through science: Why public engagement needs to move upstream. London: Demos.